BEBR

FACULTY WORKING PAPER NO. 1021

The Use of Linear Approximation to Nonlinear Regression Analysis

Anil K. Bera

College of Commerce and Business Administration
Bureau of Economic and Business Research
University of Illinois, Urbana-Champaign

BEBR

FACULTY WORKING PAPER NO. 1021

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

February 1984

The Use of Linear Approximation to
Nonlinear Regression Analysis

Anil K. Bera, Assistant Professor
Economics Department

ABSTRACT

Very often applied econometricians carry out their investigations by assuming linearity when a nonlinear model would be more appropriate. This renders the OLS estimates both biased and inconsistent. In this paper, we obtain a simple upper bound for the inconsistency. We also suggest a test for nonlinearity.

THE USE OF LINEAR APPROXIMATION TO NONLINEAR REGRESSION ANALYSIS

A. K. BERA

Department of Economics, University of Illinois

I. INTRODUCTION

We consider the following nonlinear regression equation

$$y_i = f(x_i, \alpha) + u_i, \qquad i = 1,\ldots,n , \tag{1}$$

where $y_i \in \mathbb{R}$ and $x_i \in \chi \subset \mathbb{R}^K$ are observations on dependent and fixed independent variables, $f(\cdot)$ is the response function which is assumed to be differentiable with respect to x up to a finite order, u_i is a random variable with $E(u_i) = 0$, $E(u_i^2) = \sigma^2$, $E(u_i u_j) = 0$ for $i \neq j$ and α is the unknown parameter vector.

In practical situations one cannot expect to have a precise knowledge about $f(\cdot)$. So a common procedure is to estimate a linear version of (1) like

$$y_i = x_i'\beta + \varepsilon_i, \qquad i = 1,\ldots,n , \tag{2}$$

with usual assumptions for ε_i.

When (2) is the true model OLS applied to (2) gives consistent, unbiased and efficient estimates. None of these desirable properties survives when (1) is the true model. However, these OLS estimates of β are not entirely useless. In a recent pioneering article White [4] demonstrated that the OLS estimates of the parameters of linear approximation model are inconsistent and found bounds for inconsistency. In particular, he showed that the extent of the inconsistency depends directly on the degree of concavity of the underlying nonlinear function and the skewness and range of the independent variable, and inversely on the variance of the independent variable. Using the OLS estimates he also devised a test for functional misspecification. When $f(\cdot)$ is known, they can be successfully used to estimate α consistently for a certain class of functions, see Bera [1] and Byron and Bera [3].

This paper has two distinct aims: firstly, to obtain a simpler bound for inconsistency for the OLS estimates of β when (1) is the true model and secondly, to develop a new test for functional misspecification which we shall call a test for nonlinearity.

II. A BOUND FOR INCONSISTENCY OF OLS ESTIMATES

By taking a first order Taylor series approximation of $f(\cdot)$ around the origin, (1) can be written as

$$y_i = x_i'\beta + R_i + u_i, \qquad i = 1,\ldots,n, \tag{3}$$

where β is some function of α and R_i is the remainder term. Let $\hat{\beta}$ be the OLS estimate of β from (3) without the remainder $R = (R_1,\ldots,R_n)'$, i.e.,

$$\hat{\beta} = (X'X)^{-1}X'Y ,$$

where $X = (x_1,\ldots,x_n)'$ and $Y = (y_1,\ldots,y_n)'$. It follows that

$$\operatorname{plim} \hat{\beta} = \beta + \lim \left(\frac{X'X}{n}\right)^{-1} \left(\frac{X'R}{n}\right)$$

assuming the existence of the limit on the right hand side.

Now define a vector norm of any arbitrary vector $z(K\times 1)$ by $\|z\| = (\sum_{i=1}^{K} z_i^2)^{1/2}$, and a matrix norm of a matrix A by $\|A\| = \sup_{\|z\|\leq 1} \|Az\|$ which is equal to the square root of the maximum eigenvalue of AA^* where A^* is the conjugate transpose of A.

Using the continuity property of $\|\cdot\|$ and assuming compactness of χ, we have

$$\|\text{plim}\ \hat{\beta}-\beta\| \leq \lim \frac{1}{\sqrt{n}} \|(\frac{X'X}{n})^{-1} X'\| \cdot \lim \frac{1}{\sqrt{n}} \|R\| . \tag{4}$$

Individual terms in R are (noting that we take derivatives only with respect to the last K-1 nonconstant independent variables)

$$R_i = \frac{1}{2} (x_{i2},\ldots,x_{iK}) \begin{pmatrix} \frac{\partial^2 f}{\partial x_{i2}^2} & \cdots & \frac{\partial^2 f}{\partial x_{i2}\partial x_{iK}} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_{iK}\partial x_{i2}} & \cdots & \frac{\partial^2 f}{\partial x_{iK}^2} \end{pmatrix} \begin{pmatrix} x_{i2} \\ \vdots \\ x_{iK} \end{pmatrix}$$

$$= \frac{1}{2} \underline{x}_i' D_i \underline{x}_i \quad \text{(say)}, \qquad i = 1,\ldots, n ,$$

where D_i is evaluated at $x_i^* \in (0,\underline{x}_i)$. Using the fact that for any symmetric matrix A

$$\sup_z \frac{z'Az}{z'z} = \lambda_{max},$$

where λ_{max} is the maximum eigenvalue of A, we have

$$R_i \leq \frac{1}{2} \underline{x}_i'\underline{x}_i \lambda_i , \qquad i = 1,\ldots, n ,$$

where λ_i is the maximum eigenvalue of D_i. That is

$$R \leq \frac{1}{2}\, \text{diag}(\underline{X}\underline{X}')\lambda$$

or

$$\|R\| \leq \frac{1}{2}\, \|\text{diag}(\underline{X}\underline{X}')\| \cdot \|\lambda\|, \tag{5}$$

where $\underline{X} = (\underline{x}_1, \ldots, \underline{x}_n)'$ and $\lambda = (\lambda_1, \ldots, \lambda_n)'$.

Combining (4) and (5) we have

$$\|\text{plim}\ \hat{\beta}-\beta\| \leq \frac{1}{2} \lim \frac{1}{\sqrt{n}} \|(\frac{X'X}{n})^{-1} X'\| . \lim \|\text{diag}(\underline{X}\underline{X}')\| . \lim \frac{1}{\sqrt{n}} \|\lambda\| \ , \tag{6}$$

where the terms involving X represent variability of the independent variables and $\|\lambda\|$ gives the degree of curvature. This result is similar to White's Theorem 1. To see it more clearly, assume that there is only one independent variable. Then

$$|\text{plim}\ \hat{\beta}-\beta| < \frac{x_o^2}{2} \cdot \lim(\frac{1}{n} \sum_{i=1}^{n} x_i^2)^{-1/2} \cdot \lim \frac{1}{\sqrt{n}} \|\lambda\| \ ,$$

where x_o is the largest value of x. So the degree of inconsistency depends directly on the range of x and the amount of curvature of the underlying function and inversely on the variance of x. The difference between White's and our results is that White's bounds additionally depend on the skewness of x. However, our derivation is simpler.

If we have some knowledge about $f(\cdot)$ the bound in (6) can be made tighter with appropriate scaling of data. After scaling let the new data set be

$$\overline{X} = XS,$$

where $S = \text{diag}(1,1/d,\ldots,1/d)$ with $d > 1$. The purpose of scaling is to bring all the observations on independent variables within $(-1,1)$ interval. Then (6) can be written as

$$\|\text{plim}\ \hat{\beta}-\beta\| \leq \frac{1}{2} \lim \frac{1}{\sqrt{n}} \|(\frac{X'X}{n})^{-1}X'\| \cdot \lim\|\text{diag}(\underline{X}\underline{X}')\| \cdot \frac{1}{d} \lim \frac{1}{\sqrt{n}} \|\overline{\lambda}\| ,$$

where $\overline{\lambda} = (\overline{\lambda}_1,\ldots,\overline{\lambda}_n)'$ is the same as λ but now the matrix D_i is evaluated at $\overline{x}_i^* \in (0,\overline{\underline{x}}_i)$ with $\overline{\underline{x}}_i = (\overline{x}_{i2},\ldots,\overline{x}_{iK})'$. It is clear that $\overline{\underline{x}}_i$ is nearer to the origin compared to $\underline{x}_i$. Now if the function has less curvature around the origin, then $\|\overline{\lambda}\| < \|\lambda\|$ and the bound will become much tighter.

III. A TEST FOR NONLINEARITY

It is difficult to devise a test that is applicable to all forms of functions. Therefore, we will restrict ourselves to the class of functions in which $f(\cdot)$ is monotone in x and there is a finite (possibly zero) number of inflection points throughout the whole range of x.[1]

Following White [4, p. 155] we define a parameter vector β^* which minimizes

$$Q(\beta) = [f(x,\alpha)-X\beta + u]'[f(x,\alpha)-X\beta + u].$$

Now if there are p inflection points we partition χ into $(p+2)$ disjoint sets, i.e., find $\chi_1,\ldots,\chi_{p+2}$ such that

[1]Sometimes we have some idea about a function in terms of its inflection points, e.g., most production functions do not have any inflection point whereas growth curves like logistic and Gompertz have one inflection point.

$$\bigcup_{i=1}^{p+2} \chi_i = \chi \text{ and } \chi_i \bigcap_{i \neq j} \chi_j = \phi , \qquad i,j = 1,\ldots,p+2, \quad (7)$$

where ϕ is a null set. Then define

$$Q(\beta^{i*}) = \inf_{x \in \chi_i} \{Q(\beta^i)\}, \qquad i = 1,\ldots,p+2.$$

It is postulated that testing the null hypothesis H_o: $f(\cdot)$ is linear is equivalent to testing H_o: $\beta^{1*} = \ldots = \beta^{p+2*}$. This equivalence is quite straightforward. However, the major problem is to find out an appropriate partition of χ satisfying (7). In practice, it will be difficult to satisfy the second part of (7). Therefore, instead of trying to get a disjoint partition of χ we partition the index set $\underline{I}$ = $\{1,\ldots,n\}$.

For simplicity assume there is no inflection point, in which case we partition $\underline{I}$ into two sets only. First, we order all the observations in ascending order of y, i.e., permute $\underline{I}$ in such a way that

$$y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(n)}.$$

Then choose $\underline{I}_1 = \{1,\ldots,n_o\}$ and $\underline{I}_2 = \{n_o+1,\ldots,n\}$ such that

$$\sum_{i \in \underline{I}_1} (y_i - x_i'\hat{\beta}^1)^2 + \sum_{i \in \underline{I}_2} (y_i - x_i'\hat{\beta}^2)^2$$

is minimum for $K \leq n_o \leq n-K$. This means we are fitting two "best" linear regression lines based on n_o and $(n-n_o)$ observations respectively. This will induce a partition in the observation space χ and for our sample observations we will get a partition of the matrix X. We denote it as $X = \begin{bmatrix} X_1 \\ \cdots \\ X_2 \end{bmatrix}$. Our test statistic is based on the difference $\hat{\beta}^1 - \hat{\beta}^2$ and is defined as

$$\psi = (\hat{\beta}^1 - \hat{\beta}^2)' V^{-1} (\hat{\beta}^1 - \hat{\beta}^2),$$

where

$$V = (X_1'X_1)^{-1} V_1 (X_1'X_1)^{-1} + (X_2'X_2)^{-1} V_2 (X_2'X_2)^{-1}$$

and

$$V_\ell = \sum_{i \in I_\ell} (y_i - x_i'\hat{\beta}^\ell)^2 x_i x_i' , \qquad \ell = 1,2.$$

Using Theorem 3 of White [4, p. 156], it can be shown that under H_o, ψ follows asymptotically a chi-square (χ^2) distribution with K degrees of freedom. If ψ is larger than the tabulated $\chi^2_{\alpha,K}$ value we reject H_o at α significance level.

To assess the performance of this test we consider CES production function and its two linear approximations--the Cobb-Douglas and translog functions,

(CES) $$\ln F_i = -\alpha_1/\alpha_2 \ln(\alpha_3 L_i^{-\alpha_2} + \alpha_4 K_i^{-\alpha_2}) + u_i ,$$

(Cobb-Douglas) $$\ln F_i = \beta_1 + \beta_2 \ln L_i + \beta_3 \ln K_i + \varepsilon_i ,$$

and (translog) $$\ln F_i = \beta_1 + \beta_2 \ln L_i + \beta_3 \ln K_i + \beta_4 (\ln L_i)^2 + \beta_5 (\ln K_i)^2 + \beta_6 \ln L_i \cdot \ln K_i + \varepsilon_i, \quad i = 1,\ldots,n,$$

where F_i, L_i and K_i are output, labor and capital inputs for the i-th firm. The data was generated as it was in White [4, p. 151]. We do not make much attempt to compare our results with White's reported results. His test is based on the difference between the OLS and weighted least squares estimates and it depends on the choice of weights. Different choices of weights, as presented in his paper, might lead to conflicting decisions. Our results are reported in Table 1.

TABLE 1

PARAMETER ESTIMATES FOR TWO DATA SETS AND THE TEST STATISTICS

(Sample size = 200)

	Cobb-Douglas		Translog	
	$\hat{\beta}^1$	$\hat{\beta}^2$	$\hat{\beta}^1$	$\hat{\beta}^2$
Constant	-.14437	-.08495	-.12537	-.08207
ln L	.17689	.35433	.18645	.34766
ln K	.27423	.43049	.29165	.45836
$(\ln L)^2$			.01156	.03284
$(\ln K)^2$			-.00841	-.03780
ln L.ln K			-.12351	-.05670
Dividing point (n_o)	108		108	
Test Statistic (ψ)	210.89242		207.39276	

For both the approximations the sample was divided almost at the mid point and H_o was rejected decisively. A number of other nonlinear functions, together with their linear approximations, were tried and this test procedure was found to have very high power.

IV. SOME REMARKS

Two points should be mentioned about the bound for inconsistency. First, since the bound is in terms of vector norm it is not possible to infer about individual parameters. Second, the bound depends on unknown quantities. It would be an interesting problem to investigate whether we can estimate these quantities from the available information.

While testing nonlinearity for samples of large size, finding n_o will require a huge amount of computational work. However, use of the recursive relation for matrix inversion given in Brown et al. [2, p. 152] can reduce the computational problems considerably.

Acknowledgement. I wish to thank an anonymous referee for helpful comments on an earlier draft. Errors, of course, remain my sole responsibility.

REFERENCES

1. Anil K. Bera, "Linearised estimation of nonlinear simultaneous equation systems," Working Papers in Economics and Econometrics No. 43, Australian National University, 1981.

2. R. Brown, J. Durbin and J. Evans, "Techniques for testing the constancy of regression relationships over time," Journal of the Royal Statistical Society (Series B) 37, 149-163, 1975.

3. Ray P. Byron and Anil K. Bera, "Linearised estimation of nonlinear single equation functions," International Economic Review 24, 237-248, 1983.

4. Halbert White, "Using least squares to approximate unknown regression functions," International Economic Review 21, 149-170, 1980.